



TITLE:

# Identification of Pinus species related to historic architecture in Korea using NIR chemometric approaches

AUTHOR(S):

Hwang, Sung-Wook; Horikawa, Yoshiki; Lee, Won-Hee; Sugiyama, Junji

---

CITATION:

Hwang, Sung-Wook ...[et al]. Identification of Pinus species related to historic architecture in Korea using NIR chemometric approaches. Journal of Wood Science 2016, 62(2): 156-167

ISSUE DATE:

2016-04

URL:

<http://hdl.handle.net/2433/217921>

RIGHT:

© The Author(s) 2016. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



# Identification of *Pinus* species related to historic architecture in Korea using NIR chemometric approaches

Sung-Wook Hwang<sup>1</sup> · Yoshiki Horikawa<sup>2</sup> · Won-Hee Lee<sup>1</sup> · Junji Sugiyama<sup>3</sup>

Received: 20 November 2015 / Accepted: 15 January 2016 / Published online: 22 February 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** In the present study, we identify similar *Pinus* species that have been used as building materials for traditional architecture in Korea. Discriminant models were designed to identify wood species of *Pinus densiflora* forma *erecta* Uyeki, *P. densiflora* Sieb. et Zucc., and *P. sylvestris* L. grown in Russia and Germany using near-infrared (NIR) spectroscopy coupled with multivariate analysis. Wood block samples are more practical to use in the field for wood identification; however, the face measurements have different spectral characteristics. Thus, we also prepared tablet samples from wood powder to determine the key factors for species identification without the spectral differences resulting from the wood faces. *P. densiflora* for. *erecta* and *P. densiflora* were clearly classified from *P. sylvestris* grown in Russia and Germany with a correct prediction rate of 100 % in both sample types. The discriminant models using wood block samples exhibited good performances as the  $R_p^2$  values corresponding to NIR spectral regions of 8000–4000  $\text{cm}^{-1}$  were higher than 0.90. The discriminant models using tablet samples also showed good discriminant performance. The tablet samples reduced the spectral differences of each species in the second derivative NIR spectra. However, using the tablet

samples provided more specific species information, resulting in a more accurate classification.

**Keywords** Wood identification · NIR spectroscopy · Multivariate analysis · *Pinus densiflora* forma *erecta* Uyeki · *Pinus sylvestris* L.

## Introduction

*Pinus densiflora* distributed in Korea is divided into six types by natural habitats and tree forms [1]. *Pinus densiflora* forma *erecta* Uyeki is one of the local varieties of *P. densiflora* species, and is defined by a straight trunk and outstanding properties. It is naturally distributed in Gangwon-do and northern Gyeongsangbuk-do Provinces along the Taebaek Mountains among the *P. densiflora* species in Korea. *P. densiflora* for. *erecta* is known as Geumgangsong in Korea, and is a particularly high quality pine. Their lumbering was prohibited by ancient dynasties, and they were called Hwangjangmok (yellow heartwood pine).

In Korea, pine trees have been commonly used as a building material since the Koryo Dynasty (AD 918–1392), and they became the most dominant species from the Joseon Dynasty (AD 1392–1910) to modern times [2]. In particular, *P. densiflora* for. *erecta* was used for building material and caskets for the royal family in the Joseon Dynasty. Most of the Joseon Dynasty's architectural heritages were constructed using *P. densiflora* for. *erecta* as the primary building material.

Unfortunately, the Sungnyemun Gate, also known as Namdaemun (South Main Gate), a national treasure of South Korea, was damaged by an arson fire in 2008. The fire destroyed 10 % of the first floor and 90 % of the second floor. After the fire, the Korean government took

✉ Junji Sugiyama  
[sugiyama@rishi.kyoto-u.ac.jp](mailto:sugiyama@rishi.kyoto-u.ac.jp)

<sup>1</sup> Department of Wood and Paper Science, College of Agriculture and Life Sciences, Kyungpook National University, Daegu 41566, Korea

<sup>2</sup> Faculty of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo 183-8509, Japan

<sup>3</sup> Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto 611-0011, Japan

5 years to completely restore the Sungnyemun Gate using *P. densiflora* for. *erecta* as a building material. However, after restoration was completed, it was suspected that the used timbers were not *P. densiflora* for. *erecta* but *P. sylvestris* (Scots pine) imported from Russia. In Korea, when repairing and restoring historic architecture, the same wood species as the original should be used to preserve the authenticity following Notification No. 2009-74 of the Cultural Heritage Administration of Korea [3]. Thus, it was necessary to distinguish between *P. densiflora* for. *erecta* and *P. sylvestris*. However, species differentiation was impossible using conventional methods of microscopic analysis, since the anatomical characteristics of these two species are very similar. Therefore, the genetic characteristics of the species were compared using molecular markers. The molecular markers determined that the timber suspected as *P. sylvestris* was identified as *P. densiflora* [4]. However, if the DNA cannot be probed from wood samples, it cannot be used for wood identification, and it is very difficult to probe the genetic characteristics in harvested timbers.

In previous studies, several researchers investigated chemometric approaches combined with spectroscopy and multivariate analysis for wood identification. Lewis et al. [5] investigated wood identification using near-infrared Fourier transform Raman (NIR-FTR) spectroscopy coupled with neural computing. Brunner et al. [6] reported that cluster analysis could be used to distinguish 12 different wood species. Tsuchikawa et al. [7] developed a species classification method using NIR spectroscopy combined with Mahalanobis's generalized distance, and investigated the diffusion process of deuterium-labeled molecules using NIR spectroscopy [8]. The most widely used multivariate analysis techniques for species classification are principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA). Schimleck et al. [9] classified pine and eucalyptus species, and differentiated between samples of the same eucalyptus species grown in different habitats using PCA. Antti et al. [10] determined that a PLS model could be used to distinguish individual species from mixed wood chips including Swedish pine, Swedish spruce, and Polish pine. Pastore et al. [11] discriminated true mahogany from similar species using PLS-DA. Horikawa et al. [12] used NIR spectroscopy in combination with multivariate analysis to distinguish between *P. densiflora* and *P. thunbergii*, and Hwang et al. [13] also reported that the identification of *P. densiflora* and *P. densiflora* for. *erecta* could be possible using chemometric approaches.

The purpose of the present study was to identify similar *Pinus* species used for traditional architecture. For species identification, we applied chemometrics using NIR

spectroscopy in combination with multivariate analysis, and discussed the effect of sample pretreatment on species classification.

## Materials and methods

### Samples

For species identification, wood blocks of four different *Pinus* species were used: *P. densiflora* for. *erecta* Uyeki, *P. densiflora* Sieb. et Zucc., and *P. sylvestris* L. grown in Russia and Germany. Five samples were used for each of the four species, and each wood sample was collected from different trees. The samples of *P. densiflora* and *P. densiflora* for. *erecta* were collected from Uljin city, Gyeongsangbuk-do, Korea. The wood blocks of *P. densiflora* for. *erecta* was designated as KYOw19535, 19536, 19537, 19539, and 19540 in the Xylarium at the Research Institute for Sustainable Humanosphere, Kyoto University. Those of *P. densiflora* were designated as KYOw19542, 19543, 19544, 19545, and 19546. In addition, a part of wood blocks was milled with a rough file, from which hand pressed tablet of approximately 1 g/cm<sup>3</sup> were prepared. The samples were used to examine the effect of sample pretreatment on species identification, since NIR spectra show different spectral characteristics depending on the face of wood examined [14–16].

### Optical microscopy

Samples were prepared for optical microscopy. Radial sections approximately 20 µm thick were cut using a sliding microtome, and observed after staining with safranin. The sections were observed using a light microscope (Olympus BX51) equipped with a digital camera (Olympus DP73).

### NIR spectroscopy

NIR spectra were obtained from the cross and radial section of wood blocks and tablet samples for each sample by the NIR integrating sphere diffuse reflectance accessory of the PerkinElmer Spectrum 100 N system (PerkinElmer Co., USA). The spectra were collected for wavenumbers of 10,000–4000 cm<sup>-1</sup> at a spectral resolution of 16 cm<sup>-1</sup>, and 32 scans were averaged per scan. For wood block samples, 12 spectra were acquired on the heartwood and sapwood in cross and radial sections respectively, namely, 48 spectra were obtained per sample. For tablet samples, 12 spectra were acquired per sample. Thus, 960 and 240 spectra were collected from

wood block and tablet samples, respectively. Before the multivariate analysis, the original spectrum was second derivatized by Savitzky–Golay filter smoothing to nine points with fifth order function [17].

## Multivariate analysis

PLS-DA and PCA were performed using Unscrambler v. 9.8 software (CAMO Software, Inc., Woodbridge, NJ) applied on the spectral range of 8000–4000  $\text{cm}^{-1}$ . The species identification was carried out by a one-to-one identification for each species using PLS-DA models. Therefore, four groups were formed for identification: (A) *P. densiflora* for. *erecta* and *P. sylvestris* grown in Russia, (B) *P. densiflora* for. *erecta* and *P. sylvestris* grown in Germany, (C) *P. densiflora* and *P. sylvestris* grown in Russia, and (D) *P. densiflora* and *P. sylvestris* grown in Germany. In addition, one more group was added to identify a group of the same *Pinus* species: (E) all *P. densiflora* (*P. densiflora* for. *erecta* and *P. densiflora*) and all *P. sylvestris* (*P. sylvestris* grown in Russia and Germany). Namely, a total of five groups were considered. The dataset was divided into a calibration and prediction set. Then, 320 calibration and 160 prediction samples from the wood blocks, and 40 calibration and 20 prediction samples from the tablet samples were selected randomly for each group (Table 1). For the development of the discriminant model using PLS-DA models, we assigned a class value of +1 or −1 in the calibration set. Table 2 shows discriminant groups and the class value of each sample. The performance of the discriminant model was validated with an independent prediction set of samples, consisting of 160 and 40 spectra from wood blocks and tablet samples, respectively, for each group. The coefficient of determination for calibration ( $R_c^2$ ) and the root mean square error of calibration (RMSEC) were used to assess the calibration performance. The models were evaluated using the coefficient of determination of prediction ( $R_p^2$ ) and the root mean square error of prediction (RMSEP).

## Results and discussion

### Anatomical characteristics

*Pinus densiflora*, *P. densiflora* for. *erecta*, and *P. sylvestris* have very similar anatomical characteristics. *P. densiflora* and *P. densiflora* for. *erecta* are the same species with different habitats and tree forms. In the identification keys of these species, they have large pinoid pits, resin canals, ray tracheids, and spiral thickenings in the tracheid walls. They also have distinct growth ring boundaries, abrupt transitions from earlywood to latewood, and large resin canals with thin epithelial cells. For *P. densiflora*, the dentated walls in ray tracheids are the key anatomical characteristics from other *Pinus* species. However, *P. sylvestris* also has dentated walls in ray tracheids (Fig. 1). For these reasons, they cannot be identified by conventional microscopic methods. Therefore, we applied a chemometric approach using NIR spectroscopy combined with multivariate analysis for wood identification.

### Characteristics of NIR spectra

Figure 2 is the original NIR spectra and the second derivative spectra of *P. densiflora* for. *erecta*, *P. densiflora*, and *P. sylvestris* grown in Russia and Germany in the spectral region of 10,000–4000  $\text{cm}^{-1}$ . In the original spectra, it has been difficult to distinguish each species, because the spectral characteristics appeared almost identical. However, in the second derivative spectra, some differences were revealed in the absorption bands at 7000, 5464, 5220, 4890–4620, 4404, and 4280  $\text{cm}^{-1}$  for each species. In these absorption bands, all bands except for 5220  $\text{cm}^{-1}$  assigned to water [18], are associated with cellulose among the wood components. The absorption bands at 7000, 5464 and 4280, and 4890–4620  $\text{cm}^{-1}$  are assigned to amorphous regions in cellulose, semi- or crystalline regions in cellulose, and cellulose, respectively [8, 18, 19]. The absorption band of 4404  $\text{cm}^{-1}$  is assigned to cellulose and hemicellulose, and 5980  $\text{cm}^{-1}$  is the

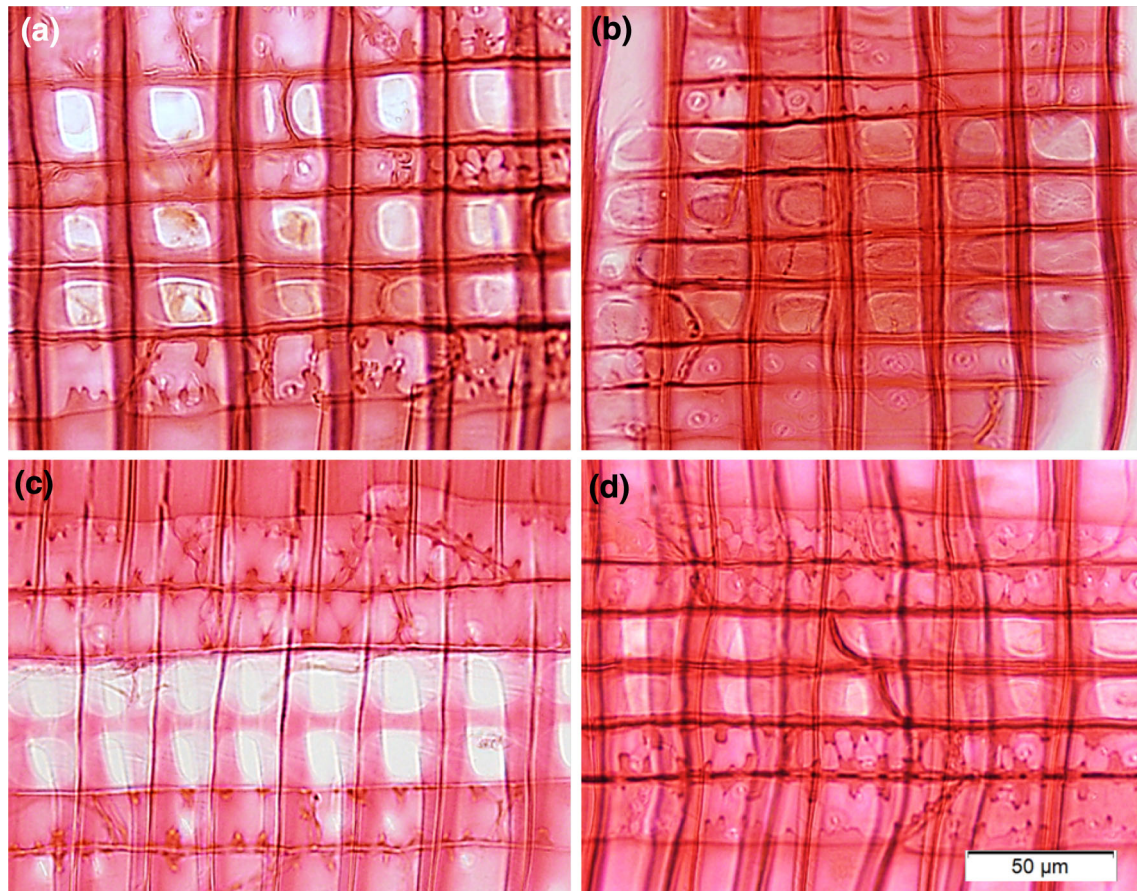
**Table 1** The number of near-infrared (NIR) spectra used for calibration and prediction sets

Species	Sample type	Calibration set	Prediction set	Total
<i>P. densiflora</i> for. <i>erecta</i>	Wood block	160	80	240
	Tablet	40	20	60
<i>P. densiflora</i>	Wood block	160	80	240
	Tablet	40	20	60
<i>P. sylvestris</i> grown in Russia	Wood block	160	80	240
	Tablet	40	20	60
<i>P. sylvestris</i> grown in Germany	Wood block	160	80	240
	Tablet	40	20	60
Total		800	400	1200



**Table 2** The wood species and class value of each group for species identification using partial least squares discriminant analysis (PLS-DA)

Groups	Species [class value]
A	<i>P. densiflora</i> for. <i>erecta</i> [+1]— <i>P. sylvestris</i> grown in Russia [−1]
B	<i>P. densiflora</i> for. <i>erecta</i> [+1]— <i>P. sylvestris</i> grown in Germany [−1]
C	<i>P. densiflora</i> [+1]— <i>P. sylvestris</i> grown in Russia [−1]
D	<i>P. densiflora</i> [+1]— <i>P. sylvestris</i> grown in Germany [−1]
E	All <i>P. densiflora</i> [+1]—all <i>P. sylvestris</i> [−1]



**Fig. 1** Optical micrographs of ray tracheids with dentate thickening in the radial section. **a** *Pinus densiflora* for. *erecta*, **b** *P. densiflora*, **c** *P. sylvestris* grown in Russia, **d** *P. sylvestris* grown in Germany

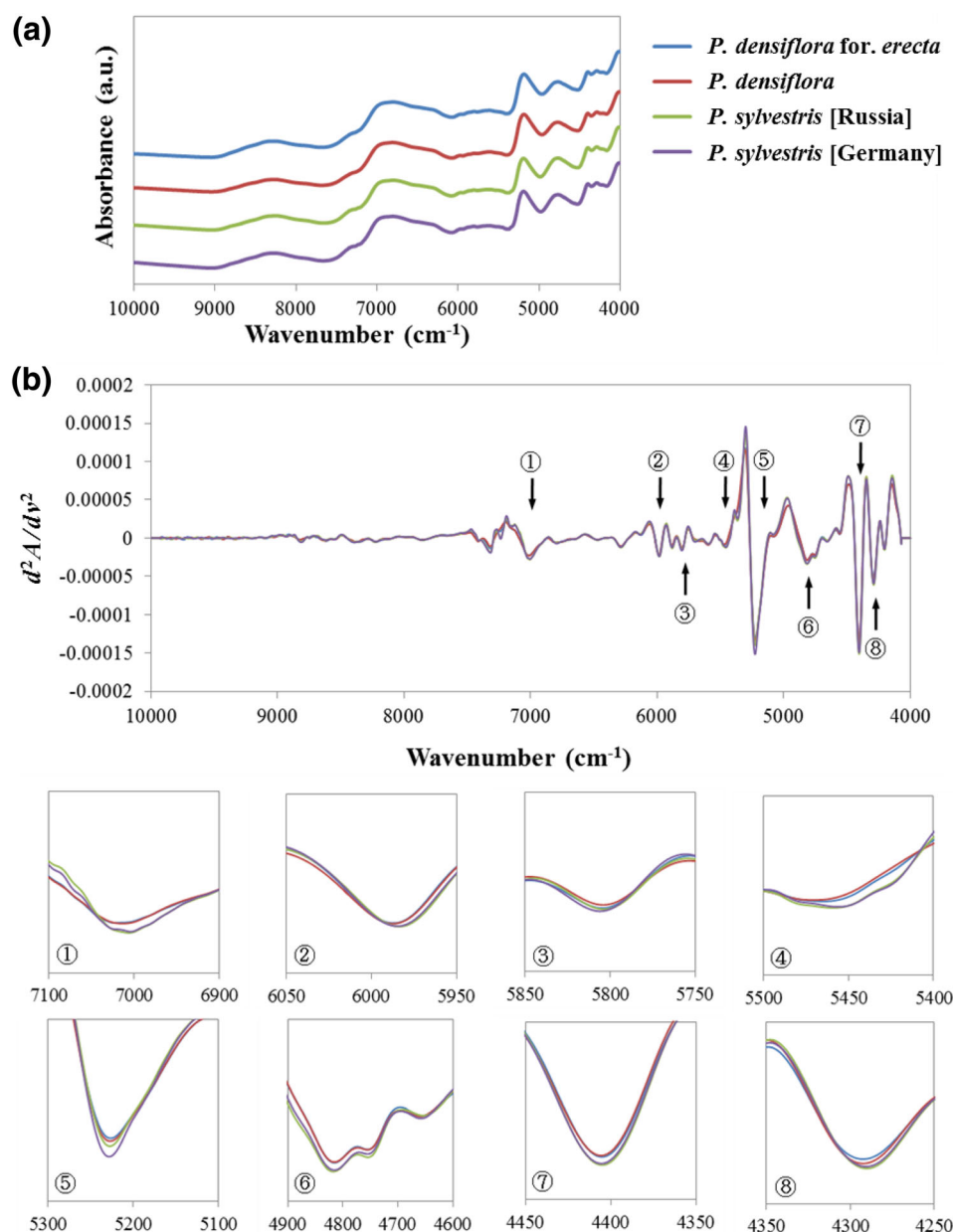
aromatic skeletal area due to lignin [18, 20]. The absorption bands of  $5800\text{ cm}^{-1}$  assigned to furanose or pyranose due to hemicellulose [18] showed similar spectral patterns for each species.

### Discriminant analysis based on block type samples

Table 3 shows the statistical summary of discriminant models for species identification based on wood block samples in the spectral region of  $8000\text{--}4000\text{ cm}^{-1}$ . All the models were relatively reliable because the  $R^2$  values of the calibration and prediction model were higher than 0.80 in all discriminant groups. RMSEP provides an

objective means of evaluating the effect of the data pre-treatment of the classification process [21]. It also can be interpreted as the average prediction error, expressed in the same units as the original response values [22]. All discriminant groups achieved 100 % accuracy using the original spectra and the second derivative spectra with a spectral region of  $8000\text{--}4000\text{ cm}^{-1}$ , except for the group E in the original spectrum. In previous studies, many researchers reported that this represents the most efficient classification performance in some specific spectral regions [12, 21, 23, 24], but the present study showed a clear classification in the spectral region of  $8000\text{--}4000\text{ cm}^{-1}$ . For the prediction model of *P.*

**Fig. 2** **a** The original near-infrared (NIR) spectra and **b** the second derivative spectra for each species based on wood block samples. *Encircled one*  $7000\text{ cm}^{-1}$  assigned to the amorphous regions in cellulose, *encircled two*  $5980\text{ cm}^{-1}$  aromatic skeletal due to lignin, *encircled three*  $5800\text{ cm}^{-1}$  furanose or pyranose due to hemicellulose, *encircled four*  $5464\text{ cm}^{-1}$  semi- or crystalline region in cellulose, *encircled five*  $5220\text{ cm}^{-1}$  water, *encircled six*  $4890\text{--}4620\text{ cm}^{-1}$  cellulose, *encircled seven*  $4404\text{ cm}^{-1}$  cellulose and hemicellulose, *encircled eight*  $4280\text{ cm}^{-1}$  semi- or crystalline regions in cellulose



*densiflora* for. *erecta* and *P. sylvestris* grown in Russia (group A), it showed the highest  $R_p^2$  and RMSEP values as  $R_p^2$  was 0.95 and RMSEP was 0.21 in the original spectra (factor 4), and  $R^2$  was 0.91 and RMSEP was 0.30 in the second derivative spectra (factor 2). The discriminant group B including *P. densiflora* for. *erecta* and *P. sylvestris* grown in Germany also showed good discriminant performance as  $R^2$  was 0.91 and RMSEP was 0.30 in the second derivative spectra (factor 2). The discriminant groups C and D including *P. densiflora* and *P. sylvestris* grown in Russia and Germany showed the same prediction performance as  $R_p^2$  was 0.90 and RMSEP was 0.32 in the second derivative spectra (factors 2 and 3).

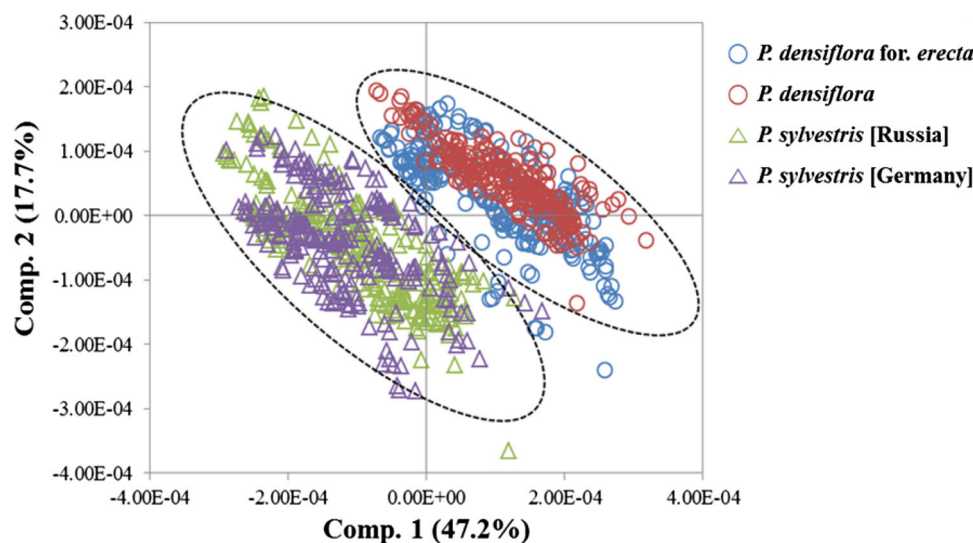
Figure 3 shows the PLS score plot of the group E based on wood block samples. In this score plot, it indicates the two clusters formed clearly. The cluster formed on the top-right side indicates mixed plots with *P. densiflora* for. *erecta* and *P. densiflora*, and lower-left indicates *P. sylvestris* grown Russia and Germany. The mixed species of each cluster are fundamentally the same species. *P. densiflora* for. *erecta* is one of the local varieties of *P. densiflora* distributed in Korea. *P. sylvestris* grown Russia and Germany are also the same species grown in different habitats. Thus, these two clusters were separated by each species as *P. densiflora* and *P. sylvestris*. From this result, these two species can be classified using the developed models on the wood block samples.

**Table 3** Statistical summary of discriminant models for species classification based on wood block samples

Groups	Spectra pretreatment	Factors	Calibration set		Prediction set		Correct prediction (%)
			$R_c^2$	RMSEC	$R_p^2$	RMSEP	
A	Original	4	0.95	0.23	0.95	0.21	100
	Second derivative	2	0.91	0.30	0.91	0.30	100
B	Original	6	0.86	0.37	0.89	0.34	100
	Second derivative	4	0.90	0.32	0.91	0.30	100
C	Original	3	0.92	0.29	0.94	0.24	100
	Second derivative	2	0.92	0.27	0.90	0.32	100
D	Original	6	0.89	0.34	0.90	0.32	100
	Second derivative	3	0.91	0.30	0.90	0.32	100
E	Original	7	0.89	0.33	0.90	0.33	99.7
	Second derivative	5	0.92	0.28	0.93	0.28	100

$R_c^2$  coefficient of determination for calibration,  $R_p^2$  coefficient of determination for prediction,  $RMSEC$  root mean square error of calibration,  $RMSEP$  root mean square error of prediction

**Fig. 3** The partial least squares (PLS) scores plot on the first two PLS components in the second derivative near-infrared (NIR) spectra with the 8000–4000  $\text{cm}^{-1}$  region based on wood block samples

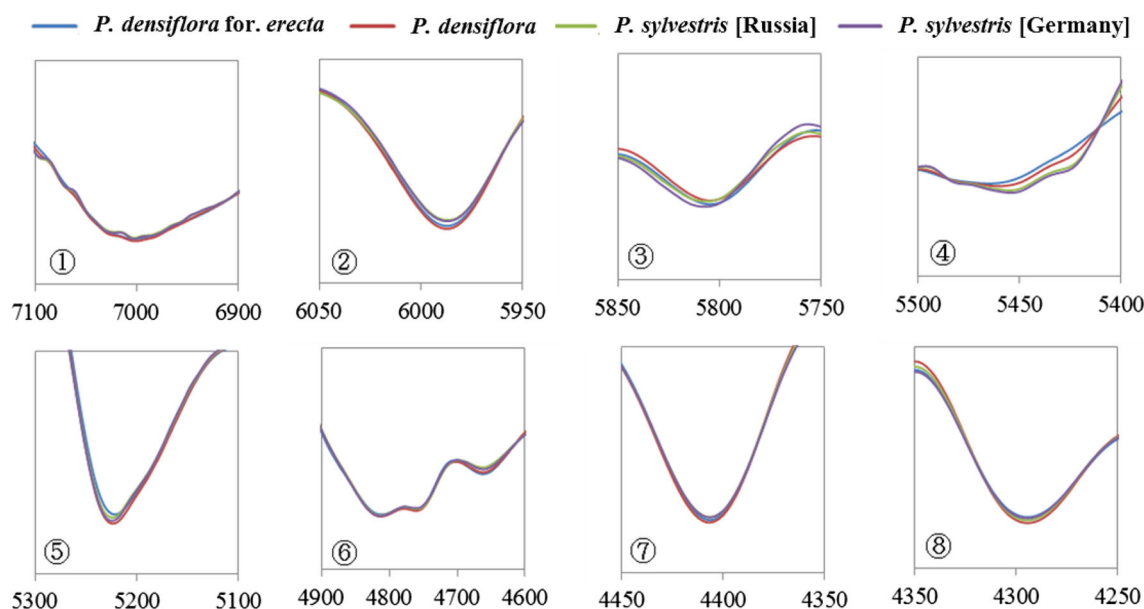


### The effect of sample pretreatment on species identification

The discriminant model was designed based on the tablet samples made from wood powder to determine the effect of sample pretreatment on species identification. Figure 4 shows several spectral regions in the second derivative NIR spectra of tablet samples. These spectral regions are the same regions referred to in Fig. 2b. As we can see in Fig. 4, the spectral differences between each species of the tablet samples were significantly reduced compared with those of the wood block samples. The spectral differences of all the absorption bands were reduced except for 5464  $\text{cm}^{-1}$ . In the absorption bands at 7000, 4890–4620, 4404, and 4280  $\text{cm}^{-1}$ , there were no spectral differences between each species. However, the absorption band at 5464  $\text{cm}^{-1}$  maintained the spectral difference in the second

derivative spectra of tablet samples. The change of NIR spectra by sample pretreatment as tablet types is possibly due to the morphological characteristics of the wood face measurements. For the wood block samples, it showed different spectral patterns on the wood surface; the absorbance was higher in the cross section than the radial section. The cross section typically showed higher absorbance than the radial and tangential section. The cross section was morphologically different from the radial and tangential section, because it had many cell lumens. That is, the NIR light travelled deeper into the cross section than into the radial and tangential section, with less reflectance in the cross section [14–16]. For the tablet sample, it had a relatively uniform surface without morphological differences on the face measurements. Therefore, it enabled a comparison of chemical components without the spectral differences of the face measurements.





**Fig. 4** The second derivative near-infrared (NIR) spectra of representative wood components based on tablet samples. *encircled one* 7000  $\text{cm}^{-1}$  assigned to the amorphous regions in cellulose, *encircled two* 5980  $\text{cm}^{-1}$  aromatic skeletal due to lignin, *encircled three* 5800  $\text{cm}^{-1}$  furanose or pyranose due to hemicellulose, *encircled four*

5464  $\text{cm}^{-1}$  semi- or crystalline region in cellulose, *encircled five* 5220  $\text{cm}^{-1}$  water, *encircled six* 4890–4620  $\text{cm}^{-1}$  cellulose, *encircled seven* 4404  $\text{cm}^{-1}$  cellulose and hemicellulose, *encircled eight* 4280  $\text{cm}^{-1}$  semi- or crystalline regions in cellulose

Figure 5 is a two-dimensional scatter plot of scores for two specific factors from the PCA on the wood block and tablet samples. The scores of the radial and cross section are roughly separated into their own clusters. As mentioned above, this is may be due to the morphological differences between the cross and radial section in the wood block samples. In the score plots of all of the species except for *P. sylvestris* grown in Russia, the scores of the tablet samples are more closely distributed in the radial section. From this result, we can assume that the tablet samples have similar morphological characteristics with the radial section than the cross section. Via et al. [25] investigated the effect of radial versus random orientation of flakes on the density prediction during composite processing. They reported that random oriented flakes by oriented strand board (OSB) processing exhibited increased error in the prediction of flake density. Defo et al. [16] reported that tangential sections showed poorer performance than cross and radial sections in the prediction model for green moisture content of red oak timber. Therefore, control of the measuring face is important to determine the precision of prediction models.

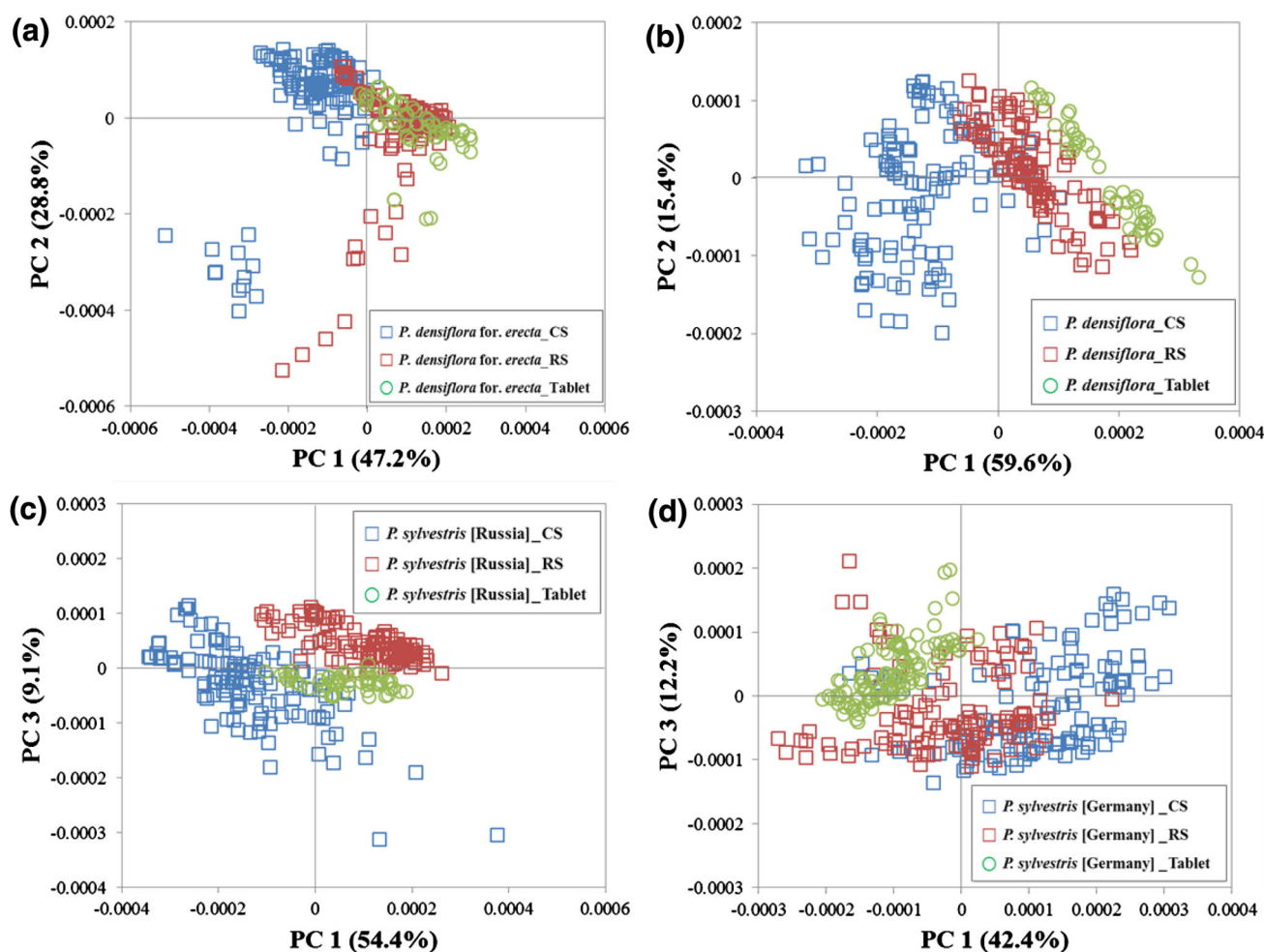
For the absorption band at 5464  $\text{cm}^{-1}$  in the second derivative spectra of tablet samples, it maintains the spectral differences between each species as in the wood block samples. This absorption band is an important region for species identification not related to the morphological differences of wood faces. Therefore, the absorption band

at 5464  $\text{cm}^{-1}$  assigned to semi- or crystalline region in cellulose is considered a key region and chemical component for species identification between *P. densiflora* and *P. sylvestris*.

Table 4 is the statistical summary of discriminant models for tablet samples. In these models, we expected poorer discriminant performance than that observed in the models for wood block samples, because the tablet samples reduced spectral differences in the second derivative spectra. Contrary to expectation, however, it also showed very high discriminant performances with correct prediction rates of 100 %. Even the discriminant group B indicated better performance than the model using wood blocks. This group, including *P. densiflora* for. *erecta* and *P. sylvestris* grown in Germany, showed the best prediction performance with an  $R_p^2$  value of 0.95 and RMSEP value of 0.21 in the second derivative spectra with factor 2.

Figure 6 is a two-dimensional scatter plot of scores for the first two factors from PLS-DA of group E based on tablet samples. The score plot of the PLS-DA model using tablet samples was separated into four clusters by each species, whereas only two clusters were formed in wood block samples. *P. densiflora* for. *erecta* and *P. densiflora* each formed a cluster on the right side based on the value 0 of component 1, and *P. sylvestris* grown in Russia and Germany formed clusters on the left side, respectively. The discriminant group B including *P. densiflora* for. *erecta* and *P. sylvestris* grown in Germany showed the farthest





**Fig. 5** The principle component analysis (PCA) scores plot on the specific two principal components in the second derivative near-infrared spectra with the 8000–4000  $\text{cm}^{-1}$  region based on wood

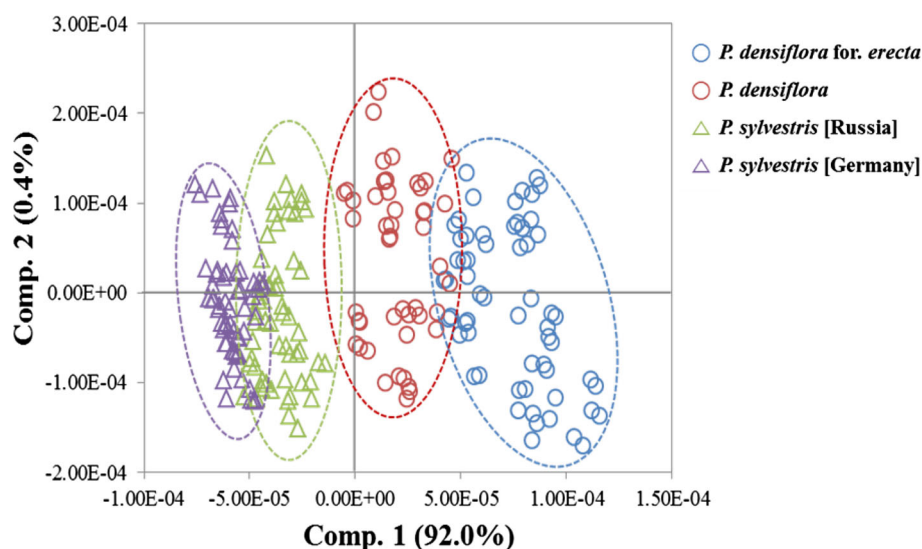
block and tablet samples. **a** *P. densiflora* for. *erecta*, **b** *P. densiflora*, **c** *P. sylvestris* grown in Russia, **d** *P. sylvestris* grown in Germany, CS cross section, RS radial section

**Table 4** Statistical summary of discriminant models for species classification based on tablet samples

Groups	Spectra pretreatment	Factors	Calibration set		Prediction set		Correct prediction (%)
			$R_c^2$	RMSEC	$R_p^2$	RMSEP	
A	Original	6	0.94	0.25	0.93	0.27	100
	Second derivative	2	0.93	0.26	0.95	0.24	100
B	Original	7	0.95	0.21	0.95	0.22	100
	Second derivative	2	0.96	0.20	0.95	0.21	100
C	Original	7	0.88	0.35	0.87	0.36	100
	Second derivative	3	0.89	0.33	0.92	0.28	100
D	Original	6	0.92	0.28	0.88	0.34	100
	Second derivative	2	0.88	0.34	0.93	0.26	100
E	Original	9	0.90	0.31	0.94	0.31	100
	Second derivative	4	0.89	0.34	0.89	0.37	100

$R_c^2$  coefficient of determination for calibration,  $R_p^2$  coefficient of determination for prediction, RMSEC root mean square error of calibration, RMSEP root mean square error of prediction

**Fig. 6** The partial least squares (PLS) scores plot on the first two PLS components in the second derivative near-infrared (NIR) spectra with the 8000–4000  $\text{cm}^{-1}$  region based on tablet samples



distance between each cluster, this group also exhibited the best discriminant performance. From these results, we found that using the tablet type samples in the development process for the discriminant model provided more specific species information, resulting in a more accurate classification. For the field application, however, the wood block sample may be more suitable because it can be applied without destruction.

### Factors affecting species identification

The discriminant models were designed to determine the factors affecting species identification based on wood block and tablet samples. From the developed models using both sample types, we determined whether species could be identified, and which components affect species identification regardless of the sample pretreatment.

Table 5 shows the statistical summary of discriminant models based on wood block and tablet samples. Generally, the models using wood block and tablet samples together showed poorer discrimination than those of wood block or tablet samples only. The number of factors and values of RMSE increased, and  $R^2$  decreased. Nevertheless, the correct prediction rates still showed high values. Discriminant group B showed the best discriminant performance in the second derivative spectra with an RMSEP value of 0.38, and  $R_p^2$  value of 0.86 with correct prediction rate of 100 %. Figure 7 shows the histograms of each discriminant model based on wood block and tablet samples in the second derivative spectra with the spectral region of 8000–4000  $\text{cm}^{-1}$ . In this region, the species identification of all discriminant groups was possible regardless of sample pretreatment.

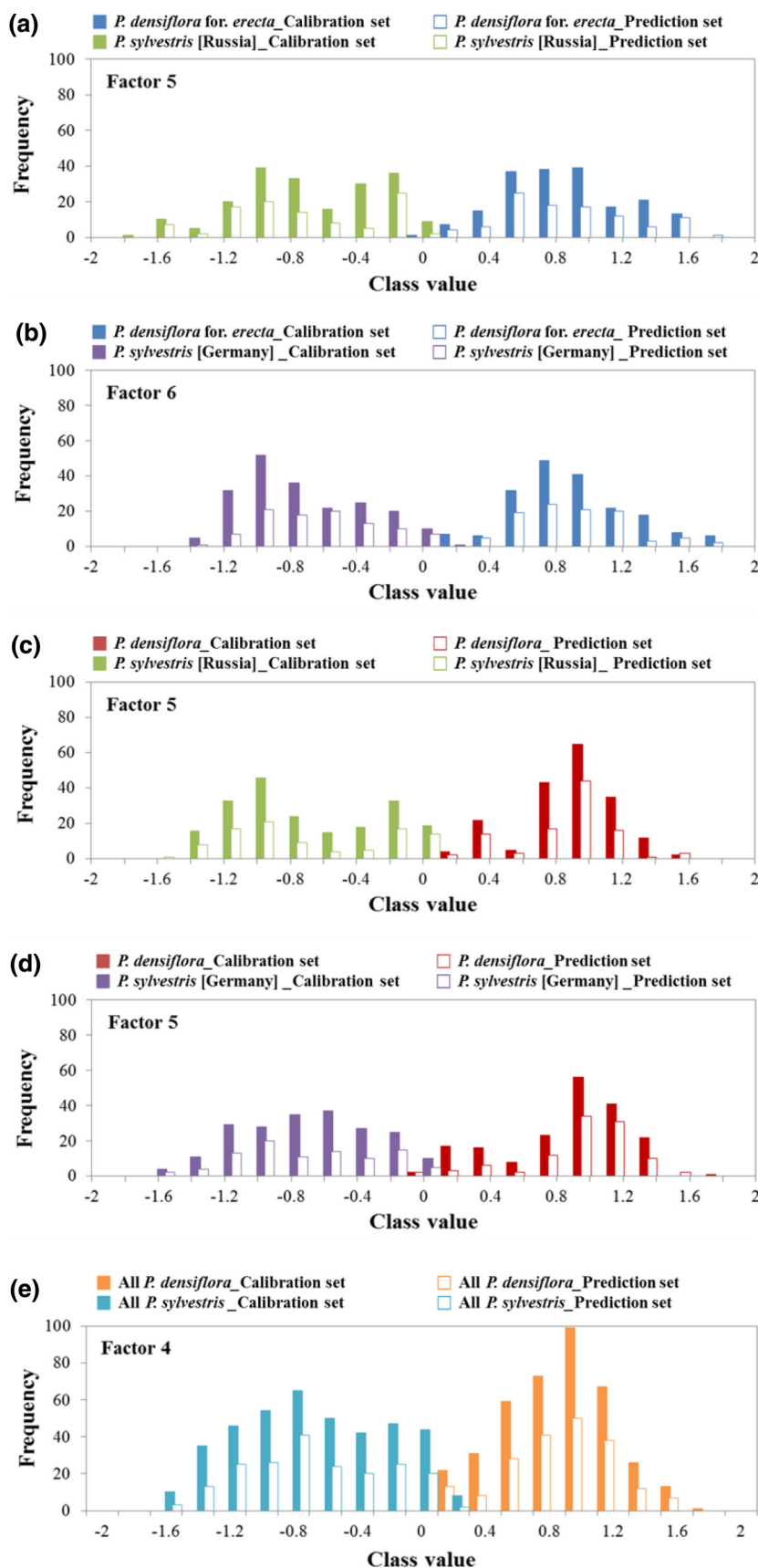
We analyzed the regression coefficients and the second derivative spectra to determine the factors affecting species

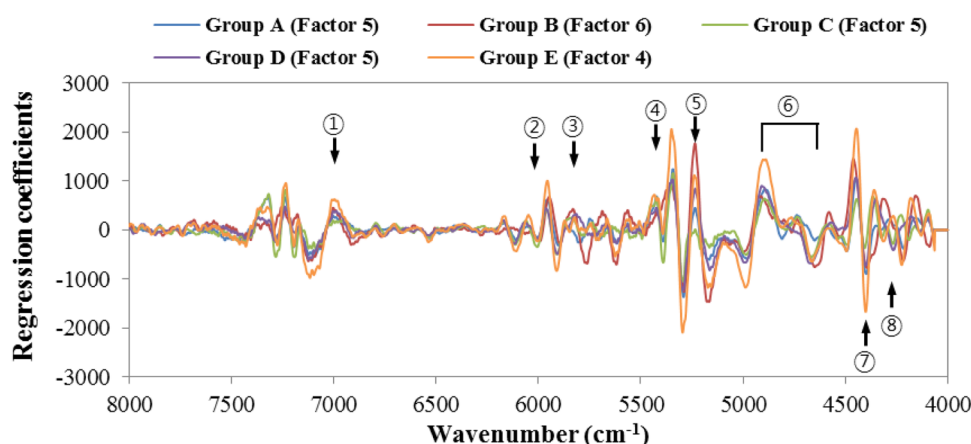
**Table 5** Statistical summary of discriminant models for species classification based on wood block and tablet samples

Groups	Spectra pretreatment	Factors	Calibration set		Prediction set		Correct prediction (%)
			$R_c^2$	RMSEC	$R_p^2$	RMSEP	
A	Original	7	0.77	0.48	0.78	0.46	99.0
	Second derivative	5	0.80	0.44	0.80	0.44	100
B	Original	7	0.78	0.46	0.81	0.44	99.5
	Second derivative	6	0.85	0.39	0.86	0.38	100
C	Original	7	0.82	0.42	0.79	0.45	97.4
	Second derivative	5	0.83	0.42	0.81	0.44	100
D	Original	6	0.80	0.44	0.79	0.46	98.5
	Second derivative	5	0.82	0.42	0.85	0.38	99.0
E	Original	8	0.77	0.48	0.80	0.43	99.0
	Second derivative	4	0.80	0.45	0.80	0.43	99.5

$R_c^2$  coefficient of determination for calibration,  $R_p^2$  coefficient of determination for prediction, RMSEC root mean square error of calibration, RMSEP root mean square error of prediction

**Fig. 7** Histograms of class value computed by partial least squares discriminant analysis (PLS-DA) in the second derivative spectra based on wood block and tablet samples. **a** *P. densiflora* for. *erecta* and *P. sylvestris* grown in Russia in factor 5, **b** *P. densiflora* for. *erecta* and *P. densiflora* grown in Germany in factor 6, **c** *P. densiflora* and *P. sylvestris* grown in Russia in factor 5, **d** *P. densiflora* and *P. sylvestris* grown in Germany in factor 5, **e** All *P. densiflora* and all *P. sylvestris* in factor 4





**Fig. 8** The spectra of regression coefficients of partial least squares discriminant analysis (PLS-DA) models for each group in second derivative spectra based on wood block and tablet samples. Each arrow indicate the following; *encircled one* 7000  $\text{cm}^{-1}$  assigned to the amorphous regions in cellulose, *encircled two* 5980  $\text{cm}^{-1}$  aromatic skeletal due to lignin, *encircled three* 5800  $\text{cm}^{-1}$  furanose

or pyranose due to hemicellulose, *encircled four* 5464  $\text{cm}^{-1}$  semi- or crystalline region in cellulose, *encircled five* 5220  $\text{cm}^{-1}$  water, *encircled six* 4890–4620  $\text{cm}^{-1}$  cellulose, *encircled seven* 4404  $\text{cm}^{-1}$  cellulose and hemicellulose, *encircled eight* 4280  $\text{cm}^{-1}$  semi- or crystalline regions in cellulose

identification (Fig. 8). The regression coefficients of the PLS-DA models using the second derivative spectra were useful to determine important spectral regions correlating to species classification [21]. The wavenumbers with an influence on the discriminant model were 7300–7000  $\text{cm}^{-1}$  and below 6000  $\text{cm}^{-1}$  in all groups. In the NIR spectra, the spectral region of 7300–6050 corresponded to the OH overtone vibrations. The spectral region of 6050–5500  $\text{cm}^{-1}$  corresponded to the CH vibrations and the vibrations from the aromatic framework, while several combinatorial vibrations were present in the region of 5500–4000  $\text{cm}^{-1}$  [12].

The spectral regions of regression coefficients indicating high influence in all groups were 7000 (peak 1), 5464 (peak 4) and 5220–5150  $\text{cm}^{-1}$  (peak 5) assigned to an amorphous region in cellulose, or a semi- or crystalline region in cellulose and water, respectively [8, 18]. In the case of group B, it showed a high value in the absorption band of 5220  $\text{cm}^{-1}$ , which means that the water had a greater influence on the species identification between *P. densiflora* for *erecta* and *P. sylvestris* grown in Germany than other discriminant groups. In the case of the second derivative spectra of wood block samples, the absorption bands at 7000 (peak 1), 5464 (peak 4), and 5220  $\text{cm}^{-1}$  (peak 5) showed some differences between all species (Fig. 2). Except for 5464  $\text{cm}^{-1}$ , however, the spectral differences of these bands were reduced in the second derivative spectra of tablet samples (Fig. 4). The absorption band at 5464  $\text{cm}^{-1}$  was still maintained the spectral differences in the spectra based on the tablet samples. Therefore, the absorption band at 5464  $\text{cm}^{-1}$  assigned to semi- or crystalline regions in cellulose is an independent spectral region that is not affected by the morphological

differences of wood faces. Namely, it can be assumed that this is the critical band and chemical component affecting the species identification of *P. densiflora* and *P. sylvestris*.

In summary, the components of cellulose and water, especially the absorption band at 5464  $\text{cm}^{-1}$  assigned to a semi- or crystalline region in cellulose were the main factors affecting the species identification in developed models based on wood block and tablet samples. Except for the spectral regions mentioned above, the absorption bands of 5350, 4458, and 4348  $\text{cm}^{-1}$  showed high regression coefficients values, but there was no assignment to the specific wood components corresponding to the these bands in previous studies. In the future, therefore, advanced analysis is required to elucidate these unknown absorption bands.

Finally, in order to avoid misunderstanding of usefulness of NIR spectroscopy, it should be mentioned that the accuracy of prediction has to be restricted in the actual datasets used in this study. It is the percentage of misclassification after leave-one-out-validation process, so that it does not mean that the unknown sample can be classified at 100 % reliability. By increasing the number of factors in discriminant model, one can arbitrarily increase the ratio of correct prediction. In this study, therefore, the factors are selected to be minimum, consisting of significant bands in the corresponding loadings, to suppress overestimation.

## Conclusion

*Pinus densiflora* for. *erecta* and *P. densiflora* were clearly distinguished from *P. sylvestris* grown in Russia and Germany using developed discriminant models. For



discriminant models using wood block or tablet samples, these showed the correct prediction rate of 100 %. Using the tablet samples reduced the spectral differences of each species; however, it provided more specific species information than wood block samples for species classification in the score plots of PLS-DA. In developed models based on wood block and tablet samples, the components of cellulose and water were the main factors affecting species identification. The absorption band at  $5464\text{ cm}^{-1}$  assigned to the semi- or crystalline region in cellulose is considered the most critical component for the identification of *P. densiflora* and *P. sylvestris*. We confirmed that a chemometric approach combined with NIR spectroscopy and multivariate analysis can be applied for the identification of similar *Pinus* species.

**Acknowledgments** This work was conducted under MOU between the Kyungpook National University and Kyoto University and partly supported by RISH Cooperative Research (database), RISH Frontier Research and Grant-in-Aid for Scientific Research from JSPS (Grant Numbers 25252033, 22300309). SWH stayed in Kyoto University as an exchange student for 6 months, which is also appreciated.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Uyeki H (1928) On the physiognomy of *Pinus densiflora* growing in Korea and silvicultural treatment for its improvement (in Japanese). Bull Agric For Coll Suigen 3:40–41
2. Park WK, Lee KH (2007) Changes in the species of woods used for Korean ancient and historic architectures (in Korean). J Archit Hist 16(1):9–28
3. Cultural Heritage Administration (2009) General principle of repair, restoration and maintenance of historic buildings and relics (in Korean). Notification No. 2009-74 of the Cultural Heritage Administration of Korea
4. Lee JW, Hong KN, Hong YP, Lee SW, Ahn JY, Kim YM (2014) Development and application of molecular markers for forest resources conservation and genetic characterization (in Korean). Korea Forest Research Institute. Research report 14–27:81–83
5. Lewis IR, Daniel NW, Chaffin NC, Griffiths PR (1994) Raman spectrometry and neural networks for the classification of wood types-I. Spectrochim Acta, Part A, Mol Spectrosc 50:1943–1958
6. Brunner M, Eugster R, Trenka E, Bergamin-Strotz L (1996) FT-NIR spectroscopy and wood identification. Holzforschung 50:130–134
7. Tsuchikawa S, Inoue K, Noma J, Hayashi K (2003) Application of near infrared spectroscopy to wood discrimination. J Wood Sci 49:29–35
8. Tsuchikawa S, Siesler HW (2003) Near-infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. Part I: softwood. Appl Spectrosc 57(6):667–674
9. Schimleck LR, Michell AJ, Vinden P (1996) Eucalypt wood classification by NIR spectroscopy and principal components analysis. Appita J 49:319–324
10. Antti H, Sjöström M, Wallbäcks L (1996) Multivariate calibration models using NIR spectroscopy on pulp and paper industrial applications. J Chemom 10:591–603
11. Pastore TCM, Braga JWB, Coradin VTR, Magalhães WLE, Okino EYA, Camargos JAA, de Muniz GIB, Bressan OA, Davrieux F (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: discrimination of true mahogany, cedar, andiroba, and curupixa. Holzforschung 65:73–80
12. Horikawa Y, Mizuno-Tazuru S, Sugiyama J (2015) Near-infrared spectroscopy as a potential method for identification of anatomically similar Japanese diploxylons. J Wood Sci 61:251–261
13. Hwang SW, Lee WH, Horikawa Y, Sugiyama J (2015) Chemometrics approach for species identification of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki (in Korean). J Kor Wood Sci Technol 43(6):701–713
14. Tsuchikawa S, Hayashi K, Tsutsumi S (1996) Nondestructive measurement of the subsurface structure of biological material having cellular structure by using near-infrared spectroscopy. Appl Spectrosc 50:1117–1124
15. Schimleck LR, Mora C, Daniels RF (2003) Estimation of the physical wood properties of green *Pinus taeda* radial samples by near infrared spectroscopy. Can J For Res 33:2297–2305
16. Defo M, Taylor AM, Bond B (2007) Determination of moisture content and density of fresh-sawn red oak lumber by near infrared spectroscopy. For Prod J 57(5):68–72
17. Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36:1627–1639
18. Siesler HW, Ozaki Y, Kawata S, Heise HM (2002) Near-infrared spectroscopy. Wiley, Weinheim
19. Osborne BG, Fearn T, Hindle PH (eds) (1993) Practical NIR spectroscopy with applications in food and beverage analysis. Longman Scientific and Tech, Harlow
20. Schwanninger M, Rodrigues JC, Fackler K (2011) A review of band assignments in near infrared spectra of wood and wood components. J Near Infrared Spectrosc 19:287–308
21. Watanabe K, Abe H, Kataoka Y, Noshiro S (2011) Species separation of aging and degraded solid wood using near infrared spectroscopy. Jpn J Histor Bot 19:117–124
22. Esbensen KH (2000) Multivariate data analysis in practice, 4th edn. CAMO, Oslo
23. Axrup L, Markides K, Nilsson T (2000) Using miniature diode array NIR spectrometers for analysing wood chips and bark samples in motion. J Chemom 14:561–572
24. Kelley SS, Rials TG, Snell R, Groom LH, Sluiter A (2004) Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Sci Technol 38:257–276
25. Via B, McDonald T, Fulton J (2012) Nonlinear multivariate modeling of strand density from near-infrared spectra. Wood Sci Technol 46:1073–1084